Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi Vol. 3, No. 2, Desember 2018
e-ISSN 2540-7902 dan p-ISSN 2541-366X

Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier

*1st Ayu Fitria
Universitas Muslim Indonesia
Fakultas Ilmu Komputer
Makassar,Indonesia
ayuFitria2327@gmail.com

2nd Muslim
Universitas Muslim Indonesia
Fakultas Ilmu Komputer
Makassar, Indonesia
muslim@umi.ac.id

3rd Huzain Azis
Universitas Muslim Indonesia
Fakultas Ilmu Komputer
Makassar, Indonesia
huzain.azis@umi.ac.id

***Abstrak*—Kebutuhan akan referensi dalam bentuk skripsi, atau artikel ilmiah semakin meningkat. Untuk mempermudah dalam menemukan informasi skripsi atau artikel ilmiah maka diperlukan sebuah sistem pengelompokan skripsi. Sistem klasifikasi dan pencarian skripsi dengan metode *naïve bayes classifier* membantu pengguna dalam penentuan topik/kategori dan menghasilkan daftar skripsi berdasarkan urutan tingkat kemiripan. Pada penelitian ini data yang digunakan adalah data skripsi yang ada pada perpustakaan Fakultas Ilmu Komputer. Proses klasifikasi pada metode *naïve bayes classifier* menggunakan dua tahap yaitu tahap *learning* dan tahap testing. Pada tahap *learning*, digunakan empat kategori dan lima dokumen di setiap kategori. Kategori yang digunakan adalah sistem informasi, *microcontroller*, Jaringan komputer, dan sistem cerdas. Setelah menentukan kategori dan jumlah dokumen data *learning* selanjutnya di lakukan perhitungan probabilitas untuk setiap kategori. Setelah melakukan perhitungan probabilitas terhadap kategori selanjutnya di lakukan perhitungan untuk data testing. Dengan adanya penelitian ini, diharapkan dapat menjadi solusi untuk memudahkan pengguna dalam mengambil informasi yang dibutuhkan sebagai bahan referensi dari kumpulan skripsi yang telah dibuat mahasiswa Fakultas Ilmu Komputer.**

Kata kunci*— *naïve bayes; skripsi; klasifikasi

I. PENDAHULUAN

Kebutuhan konsumen terhadap informasi dalam bentuk skripsi atau artikel ilmiah semakin meningkat, sehingga pengelompokan skripsi dibutuhkan untuk mempermudah pencarian informasi. Informasi menjadi kebutuhan pokok bagi setiap orang, namun tidak semua informasi yang ada dapat menjadi kebutuhan. Pemberian label topik diharapkan membantu konsumen dalam memahami isi skripsi, tanpa harus membaca secara keseluruhan. Dalam kenyataannya, pengelompokan skripsi yang mengacu topik/kategori tertentu sulit dilakukan jika hanya mengandalkan *query* biasa.

Memiliki banyak dokumen terkadang merepotkan, terutama ketika ingin mencari dokumen yang dibutuhkan dengan cepat dan tepat. Demikian halnya ketika ingin mencari informasi mengenai klasifikasi skripsi mahasiswa di Fakultas Ilmu Komputer. Kumpulan skripsi yang tersusun baik akan menjadi sumber pengetahuan yang sangat berguna.

Mencari informasi spesifik dari kumpulan skripsi ini juga sangatlah sulit, karena tidak adanya sistem klasifikasi. Untuk itu, diperlukan suatu sistem klasifikasi sebagai sebuah sistem yang mampu mencari informasi skripsi yang relevan dari sekian banyak kumpulan skripsi.

Naïve Bayes Classifier merupakan salah satu metode untuk mengklasifikasikan data. Cara kerja dari metode *Naïve Bayes Classfier* menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh Naïve bayes adalah Teorema Bayes, yaitu teorema yang digunakan dalam statistika untuk menghitung suatu peluang, *Bayes* Optimal *Classifier* menghitung peluang dari satu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal. Proses pengelompokan atau klasifikasi dibagi menjadi dua fase yaitu *learning/training* dan *testing/classify*. Pada fase *learning*, sebagian data yang telah diketahui ketegorinya, datanya diumpankan untuk membentuk model perkiraan. Kemudian pada fase testing, model yang sudah terbentuk diuji dengan sebagian data.

Data yang digunakan dalam penelitian ini adalah data skripsi yang ada di perpustakaan Fakultas Ilmu Komputer. Perpustakaan merupakan tempat yang cukup sering dikunjungi baik hanya sekedar membaca ataupun untuk mencari referensi. Fasilitas dan kenyamanan bagi pengunjung merupakan hal yang senantiasa perlu ditingkatkan di antaranya mempermudah pengunjung dalam hal pencarian skripsi dengan memanfaatkan *software* yang ada sehingga dapat membantu para pengunjung lebih cepat mengetahui daftar skripsi serta tempat penyimpanan skripsi yang ada pada perpustakaan tersebut dengan keakuratan pengklasifikasian dokumen yang baik.

Dengan adanya penelitian ini dapat menjadi solusi untuk memudahkan pengguna dalam mengambil informasi yang dibutuhkan sebagai bahan referensi dari kumpulan skripsi yang telah dibuat mahasiswa Fakultas Ilmu Komputer

II. METODOLOGI

A. *Metode Naïve*

Salah satu tugas Data Mining adalah klasifikasi data, yaitu memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya. Salah

*Corresponding Author

satu metode dalam klasifikasi data adalah *Naïve Bayes Classifier (NBC). Naïve Bayes Classifier* merupakan salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya

Dasar dari *Naïve Bayes* yang dipakai dalam pemrograman adalah rumus *Bayes* pada persamaan (1).

P(A|B)=(P(B|A)*P(A))/P(B) (1)

Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. Pada pengaplikasiannya nanti rumus ini berubah menjadi persamaan (2).

P(Ci|D)=(P(D|Ci)*P(Ci))/P(D) (2)

Naïve Bayes Classifier atau bisa disebut sebagai *Multinomial Naïve Bayes* merupakan model penyederhanaan dari Metode *Bayes* yang cocok dalam pengklasifikasian teks atau dokumen pada persamaannya (3), (4).

VMAP=argmax P(Vj|a1,a2,.....an) (3)

$$V_{MAP} = \underset{v_j \in V}{\arg max} \frac{P(a_1,a_2,...,a_n|v_j)\,P(v_j)}{P(a_1,a_2,...,a_n)} \quad (4)$$

P(a1, a2,.....an) konstan, sehingga dapat dihilangkan menjadi persamaan (5).

$$V_{MAP} = \underset{v_j \in V}{\arg max}\; P(a_1,a_2,\ldots,a_n|v_j)\,P(v_j) \quad (5)$$

Karena P(a1 , a2,….. an | vj) sulit untuk dihitung, maka akan diasumsikan bahwa setiap kata pada dokumen tidak mempunyai keterkaitan, persamaan (6), (7), (8).

$$V_{MAP} = \underset{v_j \in V}{\arg max}\; P(v_j)\prod P\,(a_i|v_j) \quad (6)$$

Keterangan :

$$P_{(v_j)} = \frac{|docs_j|}{|Contoh|} \quad (7)$$

$$P(w_k|v_j) = \frac{n_k+1}{n+|kosakata|} \quad (8)$$

Dimana,

P(vj): Probabilitas setiap dokumen terhadap sekumpulan dokumen

P(wk|vj): Probabilitas kemunculan kta wk pada suatu dokumen dengan kateogori class vj

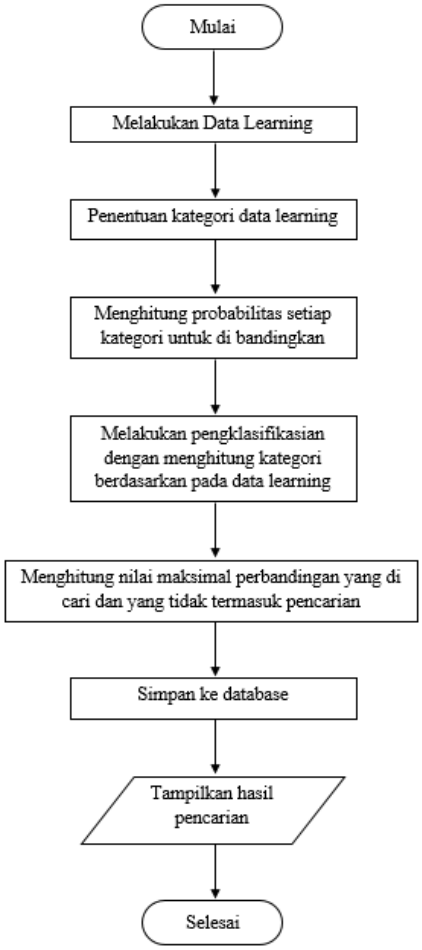
|docs| : frekuensi dokumen pada setiap kategori

|contoh| : Jumlah dokumen yang ada

Nk : Frekuensi kata ke k pada setiap kategori

Kosakata : Jumlah kata pada dokumen test

Pada persamaan (8) terdapat suatu penambahan 1 pada pembilang, hal ini dilakukan untuk mengantisipasi jika terdapat suatu kata pada dokumen uji yang tidak ada pada setiap dokumen data training. Berikut adalah alur metode Naïve Bayes Classifier untuk pengklasifikasian dokumen, Gambar 1.



Gambar. 1. Alur Metode Naïve Bayes

Berdasarkan diagram di atas dapat di lihat bahwa sebelum melakukan pengklasifikasian langkah pertama yang dilakukan adalah menentukan data learning dan kategori. Daftar judul skripsi yang akan di jadikan data testing mengacu pada kategori dan data learning yang telah di tentukan sebelumnya. Langkah terakhir yaitu menetukan tingkat kemiripan judul skripsi dengan kata kunci. Komponen penting dalam sistem klasifikasi skripsi ini adalah proses klasifikasi yang dijalankan

secara offline dan retrieval (pencarian) yang bekerja secara *online (real-time).*

B. *Cara Kerja Metode Naïve Bayes Classiffier*

Cara kerja *Naïve Bayes Classifier* melalui dua tahapan, yaitu :

1) Learning

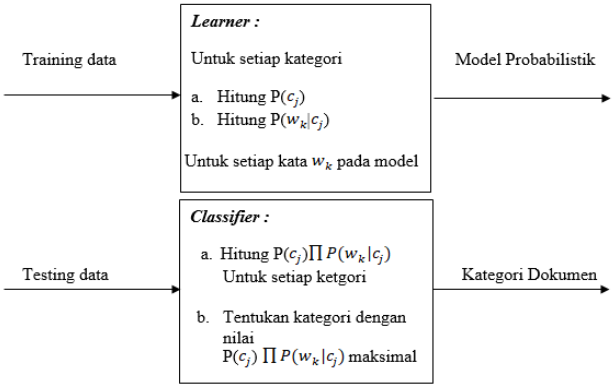
Learning (Pembelajaran) *Naïve Bayes* adalah suatu metode yang termasuk ke dalam *supervised learning*, maka akan dibutuhkan pengetahuan awal untuk dapat mengambil keputusan. Langkah-langkah :

- Bentuk *vocabulary* pada setiap dokumen data training
- Hitung probabilitas pada setiap kategori P(vj).
- Tentukan frekuensi setiap kata wk pada setiap kategori P(wk|vj)

2) Classify

Classify (Pengklasifikasian) dengan langkah-langkahnya adalah :

- Hitung P(vj)Π P(wk | vj) untuk setiap kategori.
- Tentukan kategori dengan nilai P(vj)Π P(wk | vj) maksimal, Gambar 2.



Gambar. 2. Proses Klasifikasi Dokumen Naïve Bayes Classifier

Diantara kategori yang ada biasanya dapat dikelompokkan ke dalam kategori yang lebih umum pada kasus ini hanya menggunakan empat kategori yaitu sistem informasi, microcontroller, jaringan komputer, dan sistem cerdas. Contoh data training yang digunakan ditunjukkan Tabel I.

Pada Tabel I adalah data yang di gunakan untuk data learning sebanyak 15 judul skripsi dan empat kategori. Sebelum melakukan data testing, terlebih dahulu dilakukan penetuan kategori yang akan di gunakan pada data learning. Data yang digunakan untuk data learning telah di hilangkan kata penghubungnya seperti kata "menggunakan, sebagai, untuk dan lain lain". Selanjutnya setelah menentukan data learning di lakukan perhitungan terhadap data dengan persamaan (9).

$$p(wkj|ci) = \frac{f(wkj, Ci) + 1}{f(Ci) + W} \quad (9)$$

f(wkj,Ci) adalah nilai kemunculan kata pada *wkj* pada kategori *Ci*

f(Ci) adalah jumlah keseluruhan kata pada kategori *Ci*

|W| adalah jumlah keseluruhan kata/fitur yang digunakan

Dan persamaan 10

$$p(Ci) = \frac{fd(Ci)}{|D|} \quad (10)$$

fd(Ci) adalah jumlah dokumen yang memiliki kategori *Ci*

|D| adalah jumlah seluruh training dokumen dibentuk dalam sebuah probabilitas

TABEL I. DATA LEARNING

D	Judul Skripsi	Kategori
D1	sistem(1),administasi(1),terpadu(1),berbasis(1),web(1)	sistem informasi
D2	….	sistem informasi
D3	….	sistem informasi
D19	Sistem(1), Pakar(1), Mendiagnosa(1), Psikologi(1), Penyandang(1), Autis(1), Dini(1)	Sistem cerdas
D20	Sistem(1), Pakar(1), Jenis(1), Tanaman(1), Pangan(1), Zat(1), Terkandung(1), Tanah(1)	Sistem cerdas

III. HASIL DAN PEMBAHASAN

Setelah menentukan kategori dan memilih jumlah dokumen yang akan digunakan, selanjutnya dilakukan perhitungan pada setiap dokumen atau data yang digunakan untuk setiap kategori dengan persamaan (11).

$$p(sistem\ informasi) = \frac{jumlah\ kelas\ sistem\ informasi}{jumlah\ data\ latih} \quad (11)$$

P(sistem informasi) = 5/15 = 0.333

Berdasarkan perhitungan probabilitas untuk kategori sistem informasi diperoleh nilai probabilitas yaitu 0.333, *microcontroller 0.333*, jaringan komputer 0.333, sistem cerdas 0.333, Tabel II.

TABEL II. DATA LEARNING

kategori	P(ci)	P(wkj|ci)			
		sistem informasi	microcontroller	Jaringan komputer	sistem cerdas
sistem	1/5	6/140	2/143	3/136	6/142
administrasi	1/5	2/140	1/143	1/136	1/142
terpadu	1/5	2/140	1/143	1/136	1/142
berbasis	1/5	4/140	1/143	1/136	1/142
web	1/5	3/140	1/143	1/136	1/142
…	…	…	…	…	…
zat	1/5	1/140	1/143	1/136	2/142
terkandung	1/5	1/140	1/143	1/136	2/142
tanah	1/5	1/140	1/143	1/136	2/142

Selanjutnya melakukan klasifikasi data uji menggunakan metode NBC, Tabel III menunjukkan data jurnal yang akan diuji.

TABEL III. DATA UJI

No.	Data Jurnal	kategori
D21	sistem,pendukung,keputusan, pemilihan,makanan,bergizi,balita	?

Selanjutnya masuk ketahap pengujian data jurnal, langkah awal yaitu menentukan nilai n, nc, p, dan m, berikut nilai-nilai untuk kategori "sistem informasi" ditunjukkan pada Tabel IV.

TABEL IV. NILAI DATA UJI

term	n	nc	p	m
"sistem"	20	0	0.333	7
"pendukung"	20	0	0.333	7
"keputusan"	20	0	0.333	7
"pemilihan"	20	0	0.333	7
"makanan"	20	0	0.333	7
"bergizi"	20	0	0.333	7
"balita"	20	0	0.333	7

Untuk perhitungan $P(a_i|v_j)$ yang lainnya dilakukan proses yang sama untuk setiap term disetiap kategori. Selanjutnya, dengan menggunakan persamaan 2 yaitu mencari nilai maksimal dari hasil perkalian dari nilai probabilitas.

V(sistem informasi) = 0.333*0.138*0.138*0.138*0.138* 0.138*0.138*0.138= $3{,}1738 \times 10^{-7}$

V(microcontroller) = 0.333*0.97*0.97*0.97*0.138* 0.97*0.97*0.97= 0,269

V(jaringan komputer) = 0.333*0.97*0.97*0.97*0.138* 0.97*0.97*0.97= 0.269

V(sistem cerdas) = 0.333*0.277*0.277*0.277*0.277* 0.277*0.277*0.277= 4.166×10^{-5}

Vnb = argmax(2.1738×10^{-7} | 0.269 | 0,269 | $4{,}166 \times 10^{-5}$ |)

Vnb = 4.166×10^{-5}

Sehingga didapatkan hasil bahwa jurnal D16 memiliki kategori sistem cerdas. Langkah terakhir yaitu menghitung akurasi data yang telah diuji dengan menggunakan persamaan (12).

$$Akurasi = \frac{2+0}{2+0+2+0} \times 100\% = \frac{2}{4} \times 100\% = 50\% \quad (12)$$

Dari perhitungan diatas disimpulkan bahwa hasil akurasi dari pengujian 4 data diatas menghasilkan nilai akurasi sebanyak 50%.

IV. KESIMPULAN

Metode *naïve bayes classifier* adalah salah satu metode yang digunakan untuk klasifikasi dokumen karena pada metode ini terdapat dua tahapan yaitu, data *learnig* dan data testing. Banyaknya data yang digunakan pada data learning dapat mempengaruhi hasil dari data testing.

Dalam menentukan kategori sebuah dokumen, metode *naïve bayes classifier* melakukan perhitungan data testing terhadap setiap dokumen yang ada di data *training*. Hasil dari perhitungan akan di bandingkan nilainya. Dokumen data testing akan dimasukkan pada kategori yang memiliki nilai perhitungan yang paling tinggi.

Setelah melakukan data testing, akan di lihat apakah dokumen yang di uji termasuk dalam kategori tersebut atau tidak, apabila data testing sesuai dengan kategorinya maka sistem yang dibuat sudah akurat.

Dari hasil pengujian data judul skripsi menggunakan metode *naïve bayes* dapat disimpulkan bahwa tingkat akurasi sebanyak 50%, banyaknya jumlah data yang digunakan dapat mempengaruhi hasil dari data yang diuji.

DAFTAR PUSTAKA

[1] Defianti & Jajuli 2015. Integrasi Metode Klasifikasi Dan Clustering dalam Data Mining.

[2] Ginting & Trinada. 2008. Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian Pada Aplikasi Perpustakaan.

[3] Indranandita,Susanto,Rachmat. 2008. Sistem Klasifikasi Pencarian Jurnal dengan menggunakan Metode Naïve Bayes Classifier dan Vector Space Model.

[4] Mahmudi & Widodo 2014. Klasifikasi Artikel Berita Secara Otomatis menggunakan Metode Naïve Bayes Classifier yang dimodifikasikan.

[5] Pratiwi & widodo 2017. Klasifikasi Dokumen Karya Akhir Mahasiswa Menggunakan Naïve Bayes Classifier (NBC) Berdasarkan Abstrak Karya Akhir Di Jurusan Teknik Elektro Universitas Negeri Jakarta

[6] Saiful, Taufik, Pratama 2017 . Penggunaan Naïve Bayes Classifier untuk pengelompokan pesan pada ruang percakapan maya dalam lingkungan kemahasiswaan.

[7] Rizqiyani, Mulwinda, Putri 2017. Klasifikasi Judul Buku dengan Algoritma Naive Bayes dan Pencarian Buku pada Perpustakaan Jurusan Teknik Elektro

[8] Setiawan, Astuti, Kridalaksana 2015. Klasifikasi Pencarian Buku Referensi Akademik Menggunakan Metode Naïve Bayes Classifier .

[9] Trisedya & Jais 2009. "Klasifikasi Dokumen Menggunakan Algoritma Naïve Bayes Classifier dengan Penambahan Parameter Probability Parent Category",

[10] Wijaya & Santoso 2016. Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government.